

## Exon Definition May Facilitate Splice Site Selection in RNAs with Multiple Exons

BARBARA L. ROBBERSON, GILBERT J. COTE, AND SUSAN M. BERGET\*

*Marrs McClean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030*

Received 5 July 1989/Accepted 13 September 1989

Interactions at the 3' end of the intron initiate spliceosome assembly and splice site selection in vertebrate pre-mRNAs. Multiple factors, including U1 small nuclear ribonucleoproteins (snRNPs), are involved in initial recognition at the 3' end of the intron. Experiments were designed to test the possibility that U1 snRNP interaction at the 3' end of the intron during early assembly functions to recognize and define the downstream exon and its resident 5' splice site. Splicing precursor RNAs constructed to have elongated second exons lacking 5' splice sites were deficient in spliceosome assembly and splicing activity in vitro. Similar substrates including a 5' splice site at the end of exon 2 assembled and spliced normally as long as the second exon was less than 300 nucleotides long. U2 snRNPs were required for protection of the 5' splice site terminating exon 2, suggesting direct communication during early assembly between factors binding the 3' and 5' splice sites bordering an exon. We suggest that exons are recognized and defined as units during early assembly by binding of factors to the 3' end of the intron, followed by a search for a downstream 5' splice site. In this view, only in the presence of both a 3' and a 5' splice site in the correct orientation and within 300 nucleotides of one another will stable exon complexes be formed. Concerted recognition of exons may help explain the 300-nucleotide-length maximum of vertebrate internal exons, the mechanism whereby the splicing machinery ignores cryptic sites within introns, the mechanism whereby exon skipping is normally avoided, and the phenotypes of 5' splice site mutations that inhibit splicing of neighboring introns.

Splice site selection in vertebrate pre-mRNAs occurs by recognition of consensus sequences located on the intron side of each splice site (reviewed in references 20, 32, and 52). Deletion experiments firmly indicate that the sequence elements necessary for recognition are limited to the immediate vicinity of splice sites; the bulk of the intron is dispensable (56, 57, 59, 60). Early models of splice site recognition in multi-intron RNAs suggested recognition of one site, followed by scanning through the intron to locate the second site (28, 31, 52). Experiments designed to test for intron scanning quickly indicated the implausibility of such mechanisms. First, splice sites may be mixed in chimeric RNAs or in different molecules (*trans* splicing), suggesting that the sequences characterizing individual splice sites are interchangeable and need not be connected (9, 23, 45, 53). In addition, duplication of splice sites does not result in a preference for proximal or distal sites, indicative of simple unidirectional scanning through introns (1, 27, 56, 57, 60). Cumulatively, these observations suggested that each splice site is recognized independently by specific factors and that subsequent factor interactions provide exon juxtaposition for ligation of the correct exons in the correct order.

Independent recognition of 5' and 3' splice sites, however, presents several problems for the proper orchestration of splicing, especially in transcription units with many exons. Investigation of spliceosome assembly indicates that initial substrate recognition occurs at the 3' end of the intron (4, 7, 8, 13, 18, 19, 24, 25, 43, 62, 64). In mammalian RNAs, 3' intronic sequences are not well conserved, and mutations in sequences other than the terminal AG result in little loss of function (22, 44, 47, 50). Furthermore, cryptic splice sites

are abundant in the average precursor RNA and are easily activated when normal sites are mutated (56, 57, 60). Given the lack of sequence specificity of 3' sites and the abundance of cryptic 5' and 3' sites, the mechanism whereby the splicing machinery manages to select only correct sites seems puzzling. It is also difficult to envision how neighboring exons are always ligated in correct order if little communication exists between splice sites.

Recent experiments have suggested a role for exon sequences during splice site selection (11, 14, 15, 37, 39, 40, 42, 47, 54). The phenotypes of mutations in internal 5' splice sites also suggest that splice sites are not recognized independently. Such mutations activate cryptic sites (2, 33, 56, 57, 60) or inhibit splicing of the upstream intron (33, 58). Activation of normally silent cryptic sites suggests that the splice site machinery searches a region of the precursor RNA for a consensus 5' splice site. In its absence, a neighboring best-fit splice site is selected. Utilized cryptic sites are positioned close to the correct splice site they replace; better fits to the splice site consensus sequence, but located distal to the mutation within the neighboring intron, are ignored. Thus, the site selection machinery is restricted to the immediate vicinity of the wild-type site. Exon skipping appears to occur when no usable cryptic site can be located in the vicinity of the original mutation. It should be noted that during exon skipping, the adjacent upstream intron is not removed despite the presence of wild-type 5' and 3' splice sites, suggesting that exon sequences downstream of an intron affect its recognition.

Small nuclear ribonucleoproteins (snRNPs) are prominent nuclear factors involved in splice site recognition; U1 snRNPs recognize 5' splice sites (34, 62); U2 and U5 snRNPs participate in recognition of the 3' end of the intron (5, 6, 16, 26, 41, 55). Although U1 snRNPs recognize 5'

\* Corresponding author.

splice sites, they also influence interactions at the 3' end of the intron (63, 64). We previously reported that U1 snRNPs are required for initial assembly of the spliceosome; a similar requirement is observed during yeast splicing (48). In the absence of U1 snRNP antigens or intact U1 RNA, assembly is limited to the formation of nonspecific complexes. Furthermore, other snRNPs do not stably interact with precursor RNAs when U1 snRNPs have been incapacitated. The U1 snRNP requirement for early assembly is not dependent on the presence of a 5' splice site in precursor RNA. U1 snRNP antigens are present in complexes formed with substrates lacking 5' splice sites, suggesting that U1 snRNPs participate in interactions at the 3' end of the intron. The fate of U1 snRNPs bound to the 3' end of the intron was unclear. Earlier evidence strongly indicated that such 3'-bound U1 snRNPs did not scan in a 3'-to-5' direction through the intron to select an upstream 5' splice site (27, 56, 57, 60). We were interested to test the idea that 3'-bound U1 snRNPs were destined for interactions at the neighboring downstream 5' splice site across the exon and that communication across exons occurs during initial precursor recognition. Here we report that the presence of a 5' splice site within exon 2 affects recognition and removal of the upstream intron. We suggest that factors binding 3' and 5' splice sites communicate across exons to define exons as the original units of spliceosome assembly, preceding definition of introns for catalysis.

## MATERIALS AND METHODS

**Splicing assays and substrates.** All substrates contain sequence information from the first and second exons of the adenovirus major late transcription unit. All constructs have the same last 81 nucleotides of their respective introns, 3' splice site, and first 38 nucleotides of exon 2. Substrates containing intact or truncated second exons were derived from the IVPX or MINX precursor RNA families, respectively, previously described by this laboratory (63, 64). The IVPX substrate family consists of the natural first exon from the major late transcription unit, a 231-nucleotide intron (internally deleted at the *HindIII* site) (17), the natural second exon, and over 400 nucleotides of intron 2 cloned into SP6-transcribable vectors. The MINX family of precursor RNAs was derived from a three-exon duplication construct and consists of 24 nucleotides of vector, 34 nucleotides from the second half of exon 2 (nucleotides 39 to 72), 120 nucleotides of hybrid truncated intron (the first 39 nucleotides of intron 2 connected through the *HindIII* site to the last 81 nucleotides of intron 1), and 38 nucleotides of exon 2, cloned into the SP65 polylinker at nucleotide 26.

Single-exon versions of these substrates lacking exon 1 and its 5' splice site were created by deletion between the intron-located *HindIII* site and the polylinker *EcoRI* site and are referred to as  $\Delta 5$  substrates in our earlier publications. Cleavage of IVPX DNA with *BglII*, *HpaII*, or *XhoI* produced the three length-related versions of E1 · E2 and E2 with 27, 88, or 350 nucleotides of intron 2 sequences following exon 2, respectively. A second set of substrates containing a complete second exon followed by increasingly larger downstream information was created from three exon MINX-based substrates. This family has a 94-nucleotide second exon consisting of the natural exon 2 with a 22-nucleotide insertion of SP65 polylinker at nucleotide 38. The exon is followed by 45 nucleotides of intron 2 fused directly to SP65 plasmid sequences. Truncation with *XhoI* (replacing *HindIII*

at the end of the SP65 polylinker), *HpaII*, or *PvuII* created transcription templates with 70, 180, or 286 nucleotides of sequence downstream of the 5' splice site terminating exon 2. The extending sequences contained all of the sequences terminating the MINX family of E1 · E2 $\Delta 5$  substrates that have SP65 information fused to a truncated exon 2. Thus, similar sequences were located at the 3' termini of these two families, and they are isogenic except for the region around the 5' splice site terminating exon 2 (55 nucleotides from exon 2 and 67 nucleotides of intron 2).

Cleaving MINX DNA with *BamHI*, *HpaII*, or *PvuII* produced the three length variants of E1 · E2 $\Delta 5$  and E2 $\Delta 5$  with 3, 107, or 213 nucleotides, respectively, of vector information following a truncated exon 2. Substrates with expanded second exons were created by addition of a full-length MINX DNA cassette following either the *BamHI* (in the SP65 polylinker), *HpaII* (nucleotide 132 in SP65), or *PvuII* (nucleotide 238 in SP65) site in the E1 · E2 $\Delta 5$  constructions. The additional sequences provided a 5' splice site at the end of 34 nucleotides of exon 2, a 132-nucleotide second intron modified to contain a *BglII* polylinker at the *HindIII* site, and 38 nucleotides from the beginning of exon 2. Truncation with *BglII* produced DNA templates that directed the synthesis of two-exon precursor RNAs of 322, 424, and 528 nucleotides, respectively. Truncation after the third exon (with *BamHI* or *XhoI*) produced three exon transcripts of 446, 581, and 685, respectively. Even larger second exons were produced by insertion of a 300-nucleotide *BamHI* fragment from intron 3 of the human calcitonin gene into the *BamHI* site at nucleotide 219 within the second exon in all constructs.

Truncation with *BglII* produced DNA templates that directed the synthesis of two-exon substrates of 724 and 828 nucleotides. Exon and expansion sequences are indicated in the appropriate figure legends. The expansion sequences used in all figures are identical except for short stretches of junction nucleotides necessitated by cloning. All templates were sequenced; sequences are available upon request.

Splicing reactions and analysis of complexes by RNP gel electrophoresis were performed as described previously (63, 64). Equal amounts (in counts per minute) of substrates to be compared were used. Substrates to be compared were synthesized at the same time to the same specific activity. Native RNP gels contained 3.5% acrylamide (acrylamide bisacrylamide ratio of 80:1), 0.5% agarose, 25 mM Tris, 25 mM glycine, and 10 mM EDTA (pH 8.8). Splicing reactions were mixed with a 15% volume of a concentrated stop solution and quick frozen in liquid nitrogen. Final concentration of the stop solution was 2 mg of heparin per ml–10 mM Tris chloride (pH 7.9)–2% glycerol–10 mM EDTA–0.01% bromphenol blue and xylene cyanol. Gels were run at 4°C for 4 to 6 h at 15 V/cm.

**Oligonucleotide-mediated cleavage.** Cleavage of extract U RNAs was performed as previously described (63, 64). Routinely, 0.3 to 1.0  $\mu$ g of oligonucleotide was required to cleave an individual U RNA and inhibit splicing in a standard reaction. Cleavage was monitored by examination of extract RNAs on acrylamide gels, accompanied by silver staining. Control oligonucleotides directed no observable RNA cleavage. Cleavage was for 30 min at 30°C before addition of substrate. All oligonucleotides were examined by optical density at 260 nm, gel electrophoresis and silver staining, and ability to be labeled by ATP and polynucleotide kinase. Oligonucleotide preparations were used only if they showed a single band on gel electrophoresis and kinase labeling.

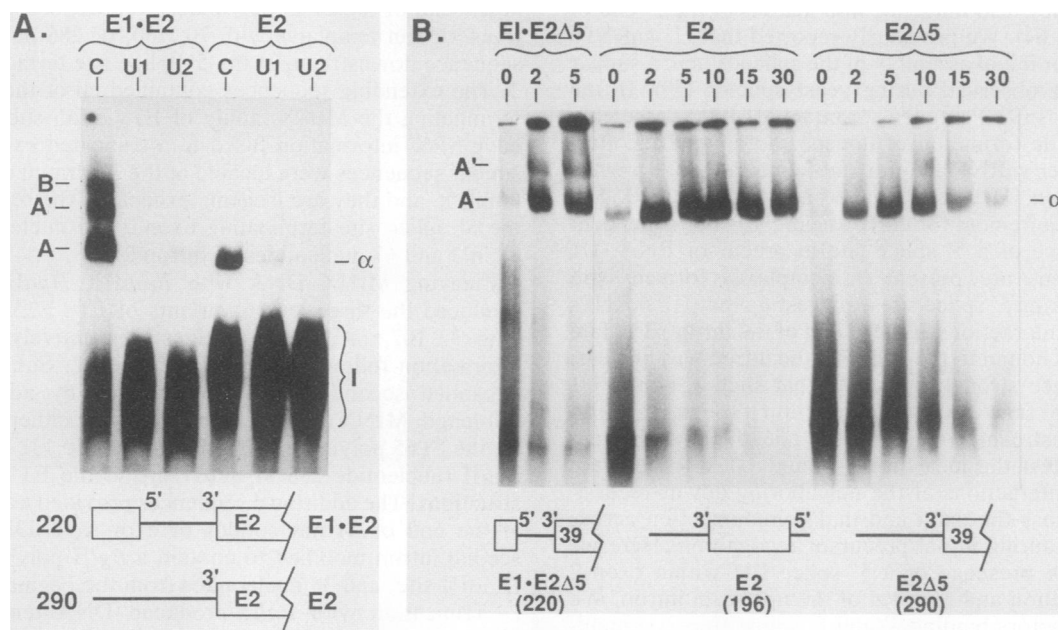


FIG. 1. RNP gel analysis of splicing reactions. Adenovirus substrates containing all or portions of exon 1, intron 1, and exon 2 were incubated in *in vitro* splicing reactions. The substrates used are diagrammed; all have 81 nucleotides of identical sequence information at the 3' end of the intron and 38 nucleotides of identical sequence beginning exon 2. Equal amounts of the substrates were used. (A) The diagrammed substrates were assembled for 15 min in splicing reactions pretreated with a control oligonucleotide (C) or oligonucleotides that direct cleavage of U1 or U2 RNA. Samples of each reaction were removed and subjected to electrophoresis on native RNP gels. The specific A, A', and B complexes assembled with wild-type substrate and the alpha complex assembled with substrates lacking exon 1 are indicated. The A, A', and B complexes correspond to the A, B, and C or alpha, beta, and gamma complexes described by other laboratories (13, 24, 30). (B) The diagrammed substrates were incubated in a standard splicing reaction for the indicated times (in minutes) before analysis on RNP gels. The smear above the alpha complex in the E2Δ5 lanes does not represent a band and is usually not observed with this substrate (see Fig. 2).

## RESULTS

**Assembly of spliceosome complexes by substrates containing a single exon is enhanced by the presence of a 5' splice site.** Simple two-exon splicing precursor RNAs assembled *in vitro* into multiple complexes that could be resolved by native RNP gel electrophoresis (Fig. 1A). We have developed a gel system (64) that resolves three complexes, A, A', and B, by using substrates with a single intron (E1-E2 in Fig. 1A). Substrates consisting of only the 3' half of the intron and a downstream second exon (E2) also assembled into a specific complex. In this case, however, only a single complex was observed, which we have termed alpha. We have previously shown that both A and alpha can be immunoprecipitated by antibodies specific for U1 snRNPs (63), suggesting that U1 snRNPs are associated with both complexes despite the absence of a 5' splice site in substrates producing alpha. No apparent cryptic 5' splice site exists in the deletion substrate producing alpha; even under extended incubation in the extract, no splicing product or intermediate RNAs are produced, suggesting that sequences at the 3' end of the intron are directing alpha assembly. Other laboratories have observed similar behavior for precursor RNAs deleted for a first exon (13, 24, 37).

U1 and U2 snRNPs are required for the formation of the A, A', and B complexes observed with substrates RNAs containing two exons (64). Formation of alpha also required intact U1 and U2 RNAs (Fig. 1A); cleavage of either U RNA arrested assembly at the formation of nonspecific complexes. The requirement for U1 RNA during the formation of complexes by using substrates that lack a 5' splice site

suggested that U1 snRNPs participate in interactions at the 3' end of the intron.

The unexpected involvement of U1 snRNPs in formation of complexes lacking 5' splice sites encouraged us to investigate whether placement of a 5' splice site downstream of the 3' splice site would have an effect on assembly. We therefore prepared substrates containing the 3' half of the intron and either an intact or a truncated exon 2 and examined their ability to direct complex formation *in vitro*. Both substrates had identical sequences at the 3' end of the intron and the beginning of the second exon.

Substrates with a complete second exon, E2, assembled rapidly (Fig. 1B). Most of the precursor RNA was present in alpha complexes by 2 min. The produced assembly was also quite stable; considerable complex remained after 30 min. In contrast, substrates with an incomplete second exon, E2Δ5, assembled less well. Considerable substrate RNA remained in nonspecific complexes after 10 min of reaction. Furthermore, by 30 min little total RNA remained, suggesting that the complex formed by using the E2Δ5 substrate afforded less protection against extract nucleases than did the complex formed on substrates containing a complete exon. In general, substrates that lack consensus sequences and are unable to form specific complexes are unstable in *in vitro* splicing extracts. Similarly, the RNA products of splicing reactions are also relatively unstable to extended incubation, presumably because of their release from stable complexes upon the completion of the reaction. The instability of the E2Δ5 RNA suggested that the alpha complex formed with incomplete second exons turns over rapidly, causing release

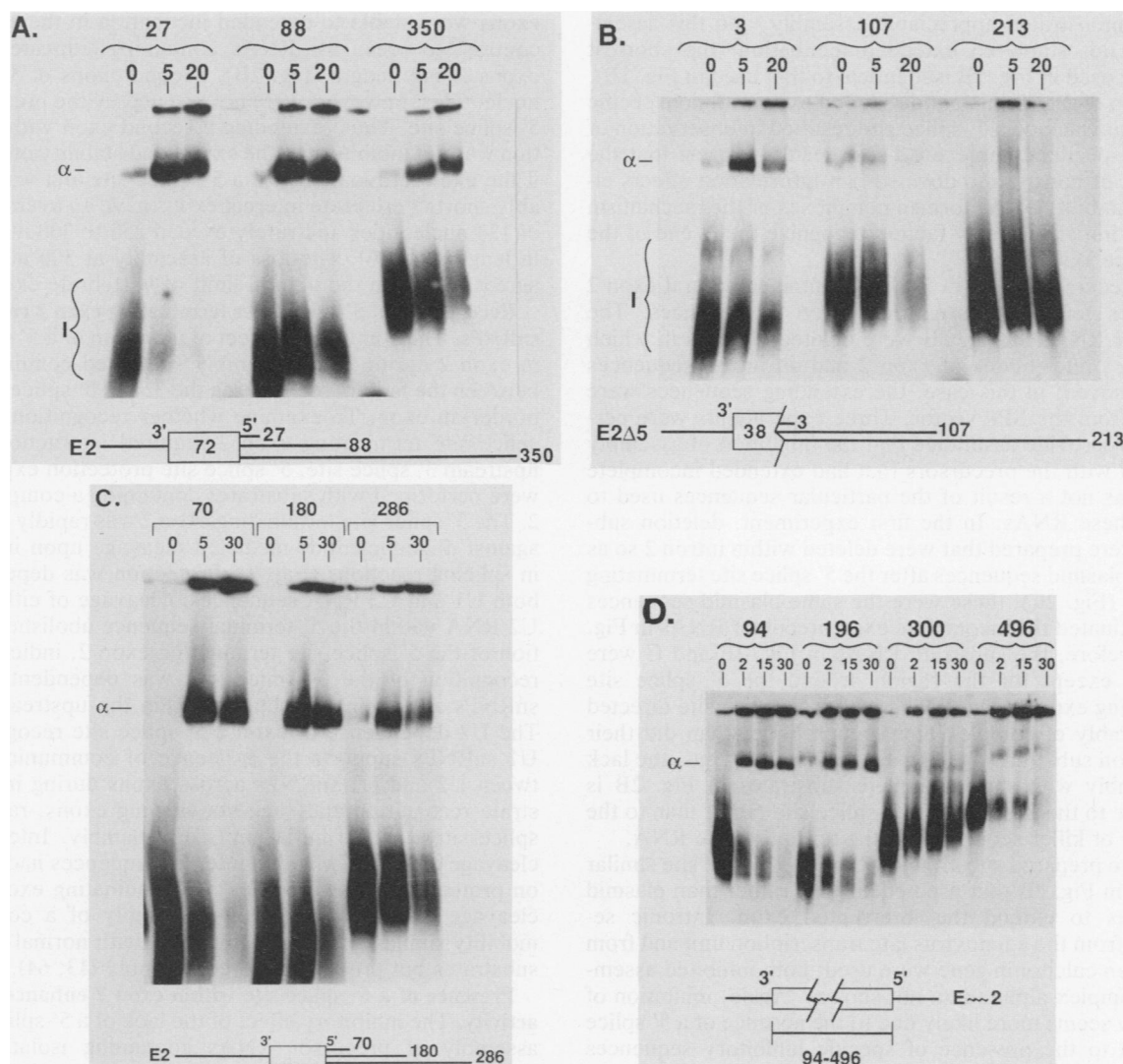


FIG. 2. Enhancement of assembly of substrates containing only exon 2 by the presence of a 5' splice site. Substrate RNAs lacking exon 1 and containing second exons of various lengths, containing or not containing a 5' splice site, were prepared. Longer exons within a set included all of the sequences present in the shorter exons. The substrates used are diagrammed. Lengths of the created second exons are indicated. The shortest version of the precursor RNAs in panels A and B were those used in Fig. 1B. (A) E2 precursor RNAs contain the natural adenovirus 2 exon 2 followed by intron 2 sequences of indicated lengths. (B) E2Δ5 precursor RNAs contain the first 38 nucleotides of adenovirus exon 2 followed by plasmid sequences of indicated lengths. (C) E2 precursor RNAs contain a 94-nucleotide (Materials and Methods) second exon followed by 67 nucleotides of intron 2 and various lengths of plasmid sequence; the precursor RNAs in panels B and C are identical except for the last 55 nucleotides of exon 2 and the first 67 nucleotides of extension. (D) E~2 precursor RNAs contain the natural adenovirus exon 2 interrupted by various lengths of plasmid sequences at nucleotide 38 of the exon; the extending plasmid sequences of the E2Δ5 precursor RNAs are the expansion sequences used in the E~2 precursor RNAs. Equal amounts of the substrates were incubated in *in vitro* splicing reactions for the indicated times before RNP gel electrophoresis. The specific alpha complexes are indicated.

of precursor RNA and concomitant exposure to nucleases. This interpretation suggests that a 5' splice site at the end of an internal exon stabilizes assembly.

To explore the sequence requirements for stabilization of alpha complexes formed with single-exon substrates, we prepared versions of exon 2 substrates that varied as to the presence and/or location of a 5' splice site (Fig. 2). The substrates that include a 5' splice site are analogous to RNAs with an exon followed by increasing lengths of intron; the substrates that lack a 5' splice site are analogous to RNAs with increasingly larger incomplete second exons. The extensions within each set were nested such that all sequences

present in the shorter substrates were also present in the longer substrates.

Equal amounts of each substrate were incubated in *in vitro* splicing reactions and sampled for alpha complex formation on native RNP gels (Fig. 2). All substrates containing a complete exon and its 5' splice site assembled complex alpha (Fig. 2A and C). Therefore, the presence of the extending nonspecific information had no effect on appearance or stability of the alpha complex. In contrast, substrates with long second exons lacking a 5' splice site within exon 2 assembled very poorly (Fig. 2B). Only the shortest substrate lacking a 5' splice site (having a second exon of 42 nucleo-

tides) demonstrated appreciable assembly, and this assembly was not stable to extended incubation (the shortest substrate used in Fig. 2B is identical to that used in Fig. 1B). Extension of the 42-nucleotide second exon with nonspecific information lacking a 5' splice site resulted in observation of only nonspecific complexes. These results suggest that the presence of nonspecific downstream information affects either the stability of the formed complexes or the mechanism whereby initial assembly factors recognize the 3' end of the intron (see Discussion).

The precursor RNAs in Fig. 2A contained natural exon 2 sequences followed by natural intron 2 sequences. The precursor RNAs in Fig. 2B were deleted variants in which the last 33 nucleotides of exon 2 and all intron sequences were removed; in this case, the extending sequences were derived from the SP6 vector. Three experiments were performed to provide assurance that the inhibition of assembly observed with the precursors that had extended incomplete exons was not a result of the particular sequences used to extend these RNAs. In the first experiment, deletion substrates were prepared that were deleted within intron 2 so as to place plasmid sequences after the 5' splice site terminating the exon (Fig. 2C); these were the same plasmid sequences that terminated the incomplete exon precursor RNAs in Fig. 2B. Therefore, the substrate RNAs in Fig. 2B and C were identical except for the region around the 5' splice site terminating exon 2. The RNAs with a 5' splice site directed the assembly of complex alpha much better than did their companion substrates lacking a 5' splice site. Thus, the lack of assembly with the incomplete substrates in Fig. 2B is likely due to the absence of a 5' splice site rather than to the presence of killer sequences at the termini of the RNA.

We also prepared substrates without a 5' splice site similar to those in Fig. 2B that used eucaryotic rather than plasmid sequences to extend the interrupted exon. Intronic sequences from the adenovirus late transcription unit and from the human calcitonin gene were used; both inhibited assembly of complex alpha (data not shown). Again, inhibition of assembly seems more likely due to the absence of a 5' splice site than to the presence of specific inhibitory sequences within the extension.

Other laboratories have reported that pyrimidine-rich exon sequences inhibit *in vitro* processing (14). The sequences used in our constructions were not pyrimidine rich. To test their inhibitory properties, however, we constructed precursor RNAs that contained a 5' splice site after the plasmid sequences used to extend the second exon (Fig. 2D). If the absence of a 5' splice site was the reason for inhibition of assembly, readdition of a 5' splice site should rescue assembly. If the extending sequences themselves were inhibitory, reintroduction of a 5' splice site should be less ameliorating. The tested precursor RNAs, designated E $\approx$ 2 in Fig. 2D, consisted of a cassette of the last 34 nucleotides of exon 2, plus the first 49 nucleotides of intron 2 added to the 3' end of the E2 $\Delta$ 5 precursor RNAs diagrammed in Fig. 2B. These precursor RNAs, therefore, contained complete second exons of 94, 196, and 300 nucleotides. A fourth precursor with an even larger second exon was created by the addition of 300 nucleotides within the exon to produce a second exon of 496 nucleotides. When the expanded exon was 94 or 196 nucleotides, stable alpha complex formation was observed. Thus, the 3 or the 107 nucleotides of extending information in the first two E2 $\Delta$ 5 extension precursor used in Fig. 2B were not inhibitory to complex formation if followed by a valid 5' splice site. Furthermore, the complexes observed with both the 94- and 196-nucleotide second

exons were stable to extended incubation in the extract, in contrast to precursor RNAs containing truncated second exons of any length (Fig. 2B). Second exons of 300 or 496 nucleotides, however, were not rescued by the presence of a 5' splice site. Thus, extending a second exon with information was not inhibitory to the extent and stability of assembly if the exon terminated with a 5' splice site and was reasonably short. Vertebrate internal exons have an average length of 134 nucleotides and rarely exceed 250 to 300 nucleotides in length (21, 36). The loss of assembly at 300 nucleotides agrees well with the natural limit of vertebrate exon size.

**Recognition of 5' splice sites terminating exon 2 requires U2 snRNPs.** The stimulatory effect of inclusion of a 5' splice site in exon 2 during early assembly suggested communication between the factors recognizing the 3' and 5' splice sites that border an exon. To examine whether recognition of the 5' splice site terminating exon 2 required interactions at the upstream 3' splice site, 5' splice site protection experiments were performed with substrates containing a complete exon 2. The 5' splice site terminating exon 2 was rapidly protected against oligonucleotide-mediated cleavage upon incubation in splicing reactions (Fig. 3). Protection was dependent on both U1 and U2 RNA sequences. Cleavage of either U1 or U2 RNA within the 5'-terminal sequence abolished protection of the 5' splice site terminating exon 2, indicating that recognition of the 5' splice site was dependent on both snRNPs and their interactions within the upstream intron. The U2 dependence of exon 2 5' splice site recognition by U1 snRNPs supports the existence of communication between U2 and U1 snRNPs across exons during initial substrate recognition and suggests viewing exons, rather than splice sites, as the initial units of assembly. Interestingly, cleavage of U2 RNA within internal sequences had no effect on protection of the 5' splice site terminating exon 2. This cleavage of U2 RNA permits assembly of a complex of mobility similar to that of complex A with normal two-exon substrates but prohibits further assembly (13, 64).

**Presence of a 5' splice site within exon 2 enhances splicing activity.** The inhibitory effect of the lack of a 5' splice site on assembly of precursor RNAs containing isolated exons suggested that removal of the 5' splice site terminating exon 2 might also inhibit splicing of two-exon substrates. Two-exon precursor RNAs were prepared that were analogous to the isolated second-exon substrates used in Fig. 2. These substrates included a second exon of 41 to 600 nucleotides with or without a 5' splice site. The first class of substrates had an intact second exon followed by up to 286 nucleotides of intron plus plasmid sequence (these substrates have 3' halves isogenic to the precursor RNAs in Fig. 2C). These E1 · E2 substrates spliced with equal efficiency regardless of the length of the extension following exon 2 (Fig. 4A). Efficiency in Fig. 4 is most easily assessed by comparing the amount of ligated E1 · E2 product (marked) or the amount of released lariat (all substrates within a single panel contain the same intron and hence the same lariat; the amount of expected lariat is related to substrate RNA length).

The second class of examined precursor RNA contained a 38-nucleotide truncated second exon without a 5' splice site followed by up to 213 nucleotides of nonspecific sequences (Fig. 4B). The precursor RNAs in Fig. 4A and B are identical except for the region containing the 5' splice site terminating exon 2. The E1 · E2 $\Delta$ 5 substrates spliced efficiently when the length of the extension was limited to 3 nucleotides (Fig. 4B). In contrast, substrates containing 107 or 213 nucleotides of extending sequences spliced poorly, as demonstrated by reduced amounts of released lariat and product E1 · E2



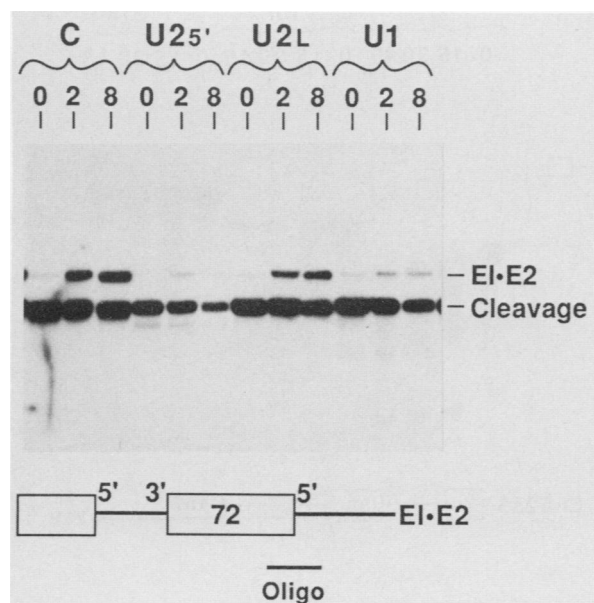


FIG. 3. Demonstration that protection of the 5' splice site in exon 2 is dependent on U2 RNA. The diagrammed E1 · E2 substrate containing the natural adenovirus exons 1 and 2 separated by a truncated intron was incubated in splicing reactions, by using extract in which U1 or U2 RNA had been cleaved with oligonucleotides complementary to the 5' portion of U2 RNA (U2<sub>5'</sub>) or U1 RNA (U1<sub>5'</sub>) or to an internal region of U2 RNA (U2<sub>L</sub>) before addition of substrate (see Materials and Methods). A control oligonucleotide not directing any observable U RNA cleavage was used to treat extract for the control (C) lane. At the indicated times, protection of 5' splice sites was monitored by addition of a complementary oligonucleotide (diagrammed) and incubation for an additional 10 min. E1 · E2 marks the position of full-length substrate, indicative of protection.

RNA. The behavior of two-exon splicing substrates in Fig. 4 was similar to that of the isolated second-exon precursor RNAs in Fig. 2. Activity was observed if a truncated second exon was quite short but not if extending downstream information was present. Again, this difference suggests that the presence of downstream information alters the association of splicing factors with the precursor RNA (see Discussion).

Extension of a truncated exon 2 may also have altered the specificity of splicing because a variety of novel lariat-exon 2 intermediates differing in gel migration (and presumably branch point) were observed with the 107-nucleotide extension. The production of unusual lariat-exon 2 species by using *in vitro* substrates with extended truncated exons has been observed previously with use of rabbit  $\beta$ -globin precursor RNAs (40). The combined loss of efficiency and specificity suggests that exon recognition plays a role in branch point selection. Thus, the presence of long second exons lacking a terminal 5' splice site disturbed the ability of the preceding intron to be correctly recognized. Notice that a perfectly good intron is present in all substrates. This behavior is reminiscent of 5' splice site mutations that result in exon skipping (33, 58).

The third class of precursor RNAs examined consisted of second exons of increasing size, from 94 to 300 nucleotides. These E1 · E2 substrates contained the same extension sequences as were present in the substrates in Fig. 4B; in this case, however, a 5' splice site was added onto the 3' end

of the extending sequences. The second exons in these constructions were identical to those used in Fig. 2 to examine the assembly of alpha complex by using precursor RNAs with isolated second exons. Readdition of a 5' splice site restored splicing to longer exons that were 94, 196, or 300 nucleotides long (Fig. 4C). The observed splicing activity with the E1 · E2 substrates was not due to activation of a cryptic 5' splice site within the extension cassettes. When a third exon was added to the constructions, both introns 2 and 3 were removed correctly (Fig. 4D). The assembly of alpha complex was normal, with precursor RNAs consisting of only a single exon when the second exon was 94 or 196 nucleotides but not 300 nucleotides long (Fig. 2). Thus, isolated exons were more susceptible to increasing exon size than were exons in two-exon substrates (see Discussion).

The ability of a 5' splice site to rescue splicing of precursor RNAs with extended exons indicates that the negative activity of the second-exon extension substrates (E1 · E2 $\Delta$ 5 precursor; Fig. 4B) was due not to the presence of inhibitory sequences within the second exon but rather to the lack of a 5' splice site. Thus, these sequences did not inhibit splicing when placed in the correct context within an exon instead of terminating one. Note that extending sequences were not inhibitory if they followed a 5' splice site, only if they replaced a 5' splice site.

The loss of splicing activity observed with two-exon substrates lacking a 5' splice site was not dependent on the nature of the sequences used to extend the exon. Substrates constructed by using procaryotic, adenovirus intronic, or calcitonin intronic sequences as the expansion cassette also failed to assemble or splice (data not shown). Because of the proven capacity of the splicing machinery to recognize cryptic splice sites, all expansion sequences used were prescreened for the absence of sequences similar to 5' splice sites. Therefore, although different sequences produced the same effect in our experiments, not all constructions of this type may yield the same phenotype if the sequences used contain cryptic consensus sequences.

We also examined the effect of very large second exons on splicing (Fig. 4E). Exons larger than 300 nucleotides inhibited the removal of the preceding intron, as indicated by a reduction in the amount of ligated product RNA. Some small amount of product, however, continued to be produced even with the larger exons. Thus, the presence of an upstream exon ameliorated the effect of a large second exon as compared with the ability of single-exon substrates to assemble complex alpha. Maximal splicing activity, however, was observed only when the exon size was 300 nucleotides or less. This size sensitivity agrees well with the observed limit on vertebrate internal exon length (21, 36).

**Assembly of two-exon substrates is enhanced by the presence of a 5' splice site in exon 2.** The splicing precursor RNAs containing second exons of various lengths with or without a 5' splice site that were used in Fig. 4 were also analyzed for assembly activity by RNP gel electrophoresis. The E1 · E2 precursor RNAs, containing a valid second exon with a 5' splice site, produced the normal pattern of A, A', and B complexes (Fig. 5A) regardless of the length or identity of the sequences past the 5' splice site. The extending sequences used in the RNAs in Fig. 5A were natural adenovirus intron sequences; similar results were observed with extending plasmid sequences (data not shown). The presence of additional downstream sequences affected neither the extent of assembly nor the gel migration of the assembled complexes. In contrast, additions of 107 or 213 nucleotides

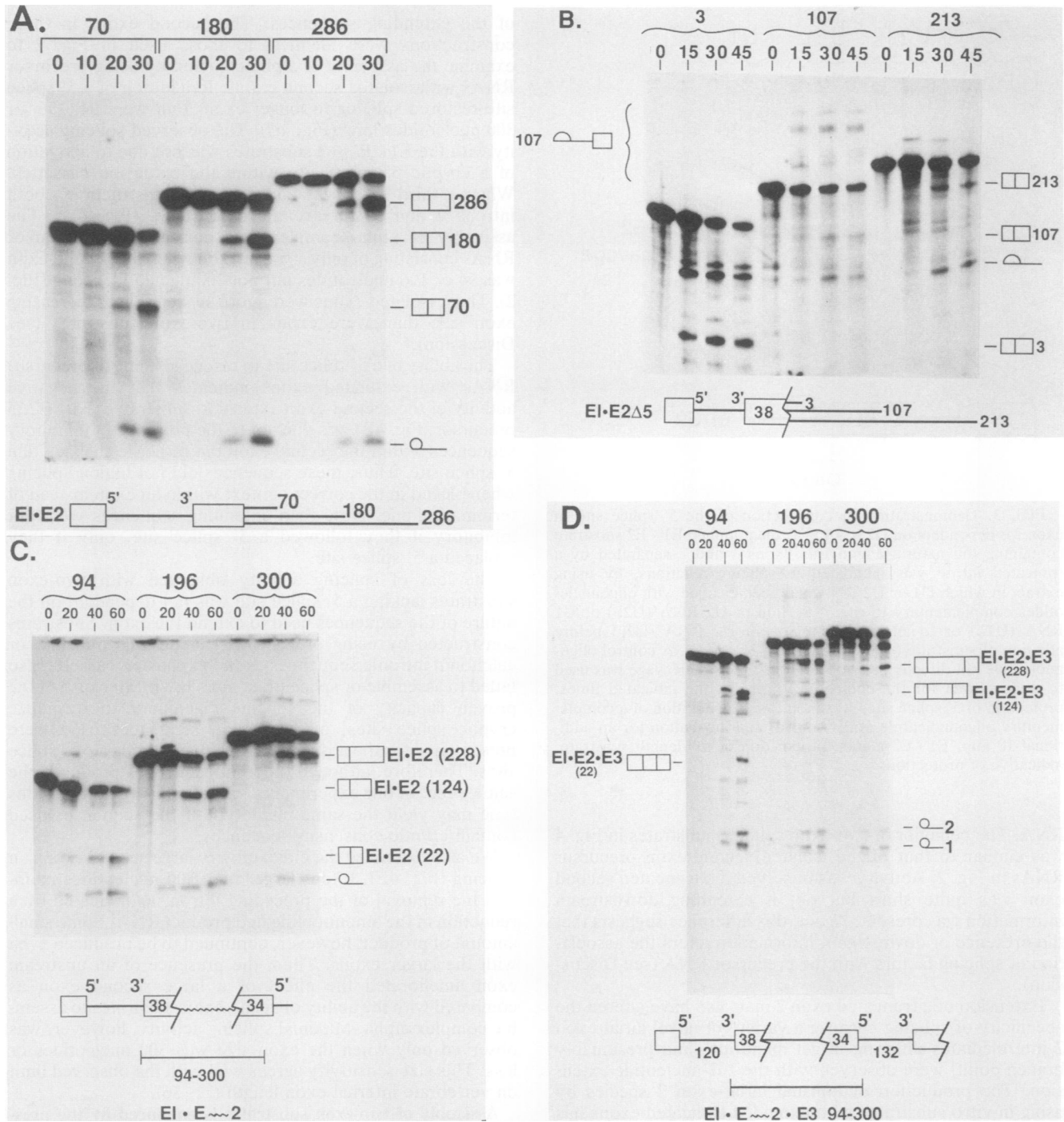


FIG. 4. Enhancement of splicing of two-exon precursor RNAs by the presence of a 5' splice site in exon 2. Splicing precursor RNAs containing an extended exon 2, with or without a 5' splice site, were prepared. The structure of each substrate RNA is diagrammed. Second-exon lengths are indicated. The second exons are identical to those used in Fig. 2. All RNAs contain the last 81 nucleotides of the natural adenovirus 2 first intron and the first 38 nucleotides of exon 2. The final products of splicing (ligated exon 1-exon 2 and released lariat) are identified. Note that whereas the precursors and products differ in length because of different exon 2 lengths, the lariats produced within the sets are identical. (A) E1·E2 precursor RNAs contain the 94-nucleotide exon 2 followed by 67 nucleotides of intron 2 and various lengths of plasmid sequence. (B) E1·E2Δ5 precursor RNAs contain the first 38 nucleotides of exon 2 followed by indicated lengths of plasmid; the precursor RNAs in panels A and B differ only by the last 55 nucleotides of exon 2 and the first 65 nucleotides of extending sequence. (C) E1·E~2 precursor RNAs contain the natural adenovirus exon 2 interrupted by various lengths of plasmid sequences at nucleotide 38 of the exon; the extending plasmid sequences of the E1·E2Δ5 precursor RNAs are the expansion sequences used in the E1·E~2 precursor RNAs. (D) E1·E~2·E3 precursor RNAs are longer versions of the E1·E~2 precursor RNAs including a second intron and a partial third exon (Materials and Methods). Introns 1 and 2 differ only by the presence of a 12-base linker in the middle of intron 2. The first 38 nucleotides of exons 2 and 3 are identical. (E) E1·E~2 precursors include those in panel C as well as two precursors containing even longer second exons created by addition of 300 nucleotides at nucleotide 38 of exon 2 of the precursors having second exons of 196 and 300 nucleotides.

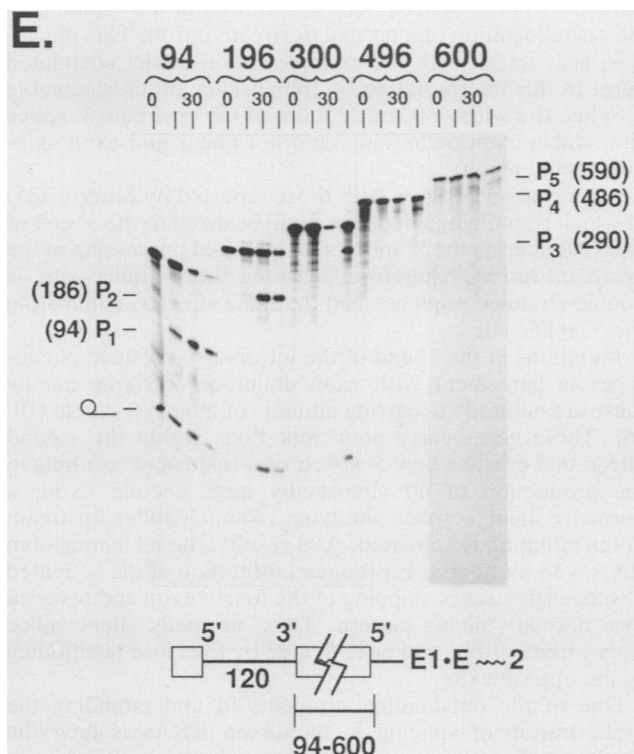


FIG. 4—Continued.

of nonspecific sequence to E1 · E2Δ5 substrates lacking a 5' splice site within exon 2 strongly inhibited assembly (Fig. 5B). A, A', and B complexes were observed only when the addition of nonspecific information was limited to 3 nucleotides. Note that the longest length (350 nucleotides) added to the permissive substrates was longer than that (213 nucleotides) added to the nonpermissive substrates. Therefore, the presence of a complete second exon with its resident 5' splice site enhanced initial spliceosome assembly.

When precursor RNAs containing expanded second exons, E1 · E2Δ5, were examined, precursors with 94, 196, and 300 nucleotides of exon all assembled well. Normal efficiency and complex mobility were observed. Therefore, extending information was inhibitory to complex formation only if it lacked a 5' splice site.

## DISCUSSION

Initial assembly of the spliceosome is dictated by interactions at the 3' end of the intron (4, 7, 8, 12, 13, 18, 19, 24, 25, 43, 46, 49). A variety of factors, including U1 snRNPs (63, 64), participate in this initial recognition step. We show here that sequences within the exon downstream of the 3' splice site facilitate initial complex formation. If these sequences contain a 5' splice site, complex formation occurs rapidly; if they do not, initial complex formation and subsequent splicing are inhibited. Therefore, when 3' intron sequences are followed by a valid exon with its resident 5' splice site, initial assembly is rapid and stable. We suggest that the exon is the initial unit of assembly (Fig. 6) and that communication between 3' splice site recognition factors (U2 and U5 snRNPs and associated proteins) and 5' splice site recognition factors (U1 snRNPs and associated proteins) functions to concertedly recognize both ends of an exon during early assembly. We term this first step exon definition. A subsequent rearrangement would be required to break contacts across exons and remake them across introns, to define introns for catalysis. RNP gel electrophoresis of splicing complexes indicates that multiple presplicing complexes are formed with simple, single-intron substrates (24, 25, 43, 64), supporting the possibility of spliceosome rearrangements and conformational changes during assembly.

We suggest, therefore, that the assembly machinery searches precursor RNA sequences hunting for 5' splice sites lying a short distance downstream of 3' intron sequences. In their presence, an exon is defined and stable assembly intermediates accumulate; in their absence, no exon is defined and stable assembly intermediates are not created. A directional search has been postulated to explain how 3' splice sites are selected downstream of initial recog-

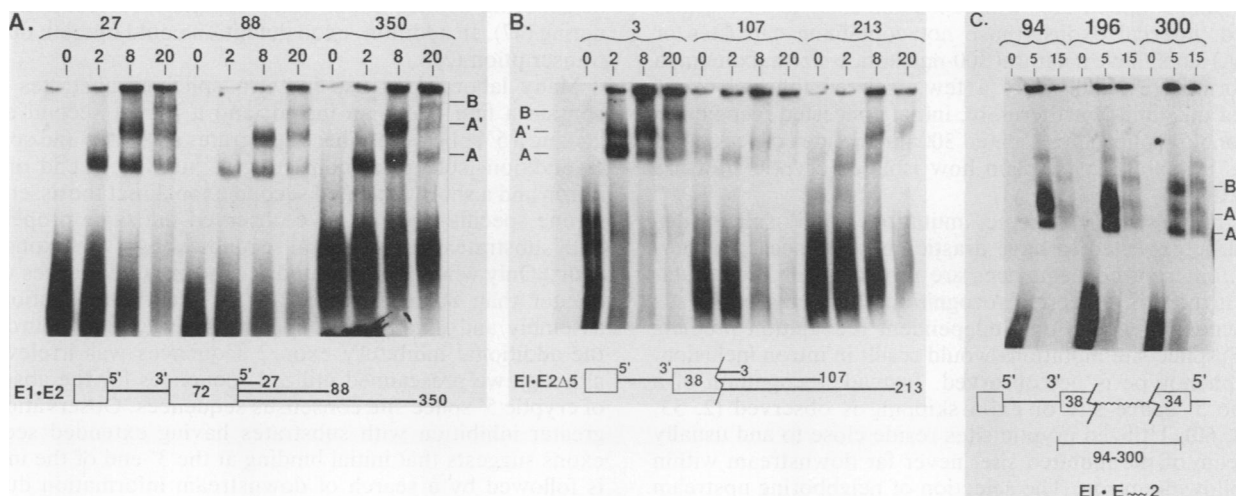


FIG. 5. Enhancement of assembly of two-exon precursor RNAs by the presence of a 5' splice site in exon 2. The reactions analyzed for assembly in Fig. 4 were sampled for assembly at the indicated times. Equal amounts of the substrates were incubated in *in vitro* splicing reactions for the indicated times and analyzed for assembly by RNP gel electrophoresis. Specific splicing complexes A, A', and B are indicated. The aberrant migration of the complexes formed with the 107- and 213-nucleotide extended second exons lacking 5' splice sites was reproducible but has not been further investigated.



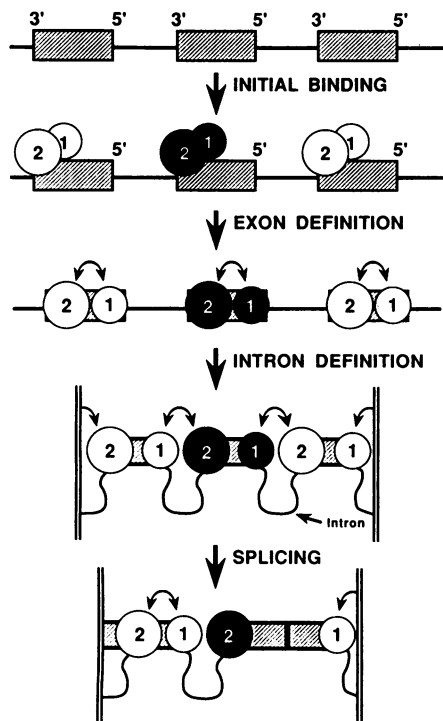


FIG. 6. Exon definition as a first step in spliceosome assembly. A postulated mechanism whereby exons are recognized as a unit during early spliceosome assembly is shown. Only an interior region of a polyexonic precursor RNA is indicated. 5'- and 3'-terminal exon recognition would require additional factors (see text). ▨, Exons. The multiple factors initially recognizing sequences at the 3' end of the intron are represented as a U2 snRNP for simplicity. A single pair of U1 and U2 snRNPs is highlighted to facilitate following a single exon.

tion of branch and pyrimidine sequences (29). Our model extends such a search through the exon. An assembly mechanism requiring concerted recognition of the 3' and 5' splice sites bounding and defining an exon would suggest that 5' and 3' splice sites will be located close to each other. Indeed, internal exons [those not containing cap sites or poly(A) sites] have a natural 300-nucleotide size maximum in vertebrates (21, 36). Only a few, isolated internal exons exceed this limit. Furthermore, initial concerted recognition of 3' and 5' splice sites located 300 nucleotides or less apart across an exon may explain how isolated cryptic sites are ignored.

In this assembly scheme, mutations of 5' splice sites would be expected to have drastic consequences. Furthermore, different consequences are predicted for such mutations if there is concerted recognition of splice sites versus independent recognition. Independent recognition predicts that 5' splice site mutations would result in intron inclusion. This phenotype is not observed. Instead, recognition of a cryptic 5' splice site or exon skipping is observed (2, 33, 56–58, 60). Utilized cryptic sites reside close to and usually upstream of the mutated site, never far downstream within the following intron. The selection of neighboring upstream cryptic sites supports a concerted splice site selection process with an inherent spacing maximum between 3' and 5' sites. The second phenotype observed for mutations of the 5' splice site of internal exons is exon skipping. Even though the intron preceding an exon with a mutated 5' splice site is

normal, it is not spliced. Thus, mutation of a 5' splice site affects recognition of a normal upstream intron. This phenotype also agrees with the exon definition model postulated here. In this interpretation, in the absence of an acceptable 5' splice site within 300 nucleotides of the upstream 3' splice site, stable exon definition could not occur and exon skipping would result.

Our results also agree with those reported by Munroe (35) in which ribo-oligonucleotides prehybridized to the 3' end of exon 2 (covering the 5' splice site) inhibited processing of the upstream intron. Therefore, masking the 5' splice site as double-stranded sequence had the same effect as eliminating the 5' splice site.

Mutations in the 3' end of the intron also produce phenotypes in agreement with exon definition. Striking among these are naturally occurring mutants of human  $\beta$ -globin (10, 56). These genes carry point mutations within the second intron that create a new 5' splice site. Instead of resulting in the production of an abnormally large second exon, a normally silent 3' splice site lying 120 nucleotides upstream of the mutation is activated. As a result, a novel hemoglobin RNA with four exons is produced. Mutation of the activated 3' splice site causes skipping of the fourth exon and reversal to a normal splicing pattern. Thus, normally silent splice sites within introns can be activated by the close positioning of the opposite site.

One of the outstanding problems in understanding the orchestration of splicing in multiexon RNAs is how the machinery functions so as to never unintentionally skip an exon. Intron-scanning mechanisms have been proposed, tested, and ruled out (27, 28, 31, 51, 56, 57, 60). The experiments designed to test intron scanning asked whether splice site selection is coupled to intron scanning. Here we suggest that splice site selection is coupled to exon searching. Experiments duplicating 5' splice sites within internal exons do show a preference for the upstream-most site (11). The activation of cryptic 5' splice sites lying upstream of mutated 5' splice sites, however, indicates that the searching mechanism defining exons is not a strict 5'-to-3' scan but instead is a searching mechanism that operates to find the best fit to consensus sequences. Once splice sites (and hence exons) are selected, scanning through introns to detect neighboring defined exons could facilitate correct exon ordering (47). In addition, exon definition could operate during transcription (2).

Many laboratories use *in vitro* splicing substrates that contain a first exon, an intron, and a partial second exon lacking a 5' splice site. These substrates assemble and splice. In addition, substrates consisting of just the 3' end of the intron and a short truncated second exon direct the assembly of one specific complex. We observed the same properties with substrates having partial second exons if the exons are short. Only when the truncated second exon sequences were greater than 80 nucleotides did we see destabilization of assembly and inhibition of splicing activity. The source of the additional inhibitory exon 2 sequences was irrelevant, although we prescanned utilized sequences for the absence of cryptic 5' splice site consensus sequences. Observation of greater inhibition with substrates having extended second exons suggests that initial binding at the 3' end of the intron is followed by a search of downstream information during exon recognition. If no 5' splice site is encountered within 300 nucleotides, no stable complex is formed. This is the situation that would occur when an isolated splice site is located within an intron. Thus, we suspect that the extended exon substrates are good models for understanding the poor

recognition of isolated cryptic splice sites by the splicing machinery.

Substrates having short truncated second exons are not normally encountered by splicing factors; therefore, the ability of these substrates to work in vitro may be a property of the in vitro system. In this interpretation, in the absence of sequences to search, such as in substrates containing short truncated second exons, the produced assembly is marginally stable. Indeed, although complexes were observed with precursor RNAs consisting of just the 3' end of the intron and a short truncated second exon, they were unstable to extended incubation in processing extract in comparison with similar substrates containing a 5' splice site (Fig. 2).

Complexes formed with substrates that consisted of the 3' end of the intron and a complete second exon with a 5' splice site were extremely stable to incubation in the extract. It would be predicted that the definition of exons in vivo should be both fast and irreversible. Failure or loss of definition of an exon would result in exon skipping. Indeed, the first complexes observed during in vitro assembly occur within 1 min of incubation in vitro and on nascent transcripts in vivo (3). Higher-order complexes involved with intron definition as depicted in Fig. 6 appear with slower kinetics in vitro, suggesting that intron definition is the rate-limiting step during spliceosome assembly.

In this communication, we report that the absence of a 5' splice site within long second exons was more inhibitory for the assembly of substrates lacking exon 1, i.e., substrates limited to exon definition, than it was to substrates with a complete first exon that could undergo both exon and intron definition (compare Fig. 2 and 4). Communication between factors bound at the ends of introns has been observed in vitro (30). It seems likely that this communication serves to directly define introns at reasonable efficiency in vitro as long as there are no sequences to search downstream of the 3' end of the intron. We suggest, therefore, that substrates having short truncated second exons function via direct intron definition and the bypassing of exon definition. We anticipate that bypassing of exon definition might require short introns. Most in vitro experiments use internally deleted introns to maximize complex formation and activity. Lengthening of introns generally reduces activity. We are presently constructing precursors with increasingly larger introns to test requirements of direct intron definition for those substrates blocked for exon definition because of large second exons.

The discussion presented above refers to internal exons only. Obviously, if exon definition occurs, additional mechanisms must operate to recognize first and last exons. It has been established that splicing of the first intron in a multi-intron RNA is more cap dependent than is removal of the second intron (38). Therefore, recognition of the first 5' splice site terminating exon 1 might be mediated by cap-recognizing factors. We have recently obtained evidence for the existence of soluble nuclear factors that are required for recognition of cap-proximal 5' splice sites but are not required for recognition of internal 5' splice sites (61; unpublished observation). Therefore, special nuclear factors may recognize the first exon in a precursor RNA.

Terminal 3' exons contain poly(A) sites and are frequently extremely long. If exons are the initial assembly unit, there must be special factors that aid in recognition of 3'-terminal exons. We have observed that adding a sequence containing a poly(A) site to a truncated second exon lacking a 5' splice site permits assembly of stable complexes (M. Niwa and

S. M. Berget, unpublished observations). We are presently examining the sequence elements within the poly(A) site that rescue assembly of substrates with extended exons. It seems plausible that the splicing and polyadenylation machinery communicate across terminal exons to direct assembly.

It should be noted that we observed a dependence of spliceosome assembly on second-exon sequences when we used a family of substrate RNAs that all contained similar intron and 3' splice site sequences. It is possible that RNAs with different sequences at the 3' end of the intron will respond differently to the presence of downstream information. Such differences could be exploited to produce differential splicing. Any factor that disrupts interactions at the 3' end of the intron, the process of exon searching, or interactions at the downstream end of the exon could theoretically cause exon skipping. Such an exon-oriented view of splice site selection directs attention to the exon as the location of *cis*-acting sequences operative during differential processing.

#### ACKNOWLEDGMENTS

This work was supported by awards from Public Health Service and the Robert A. Welch Foundation to S.M.B.

#### LITERATURE CITED

1. Aebi, M., H. Hornig, and C. Weissmann. 1987. 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved GU. *Cell* 50:237-246.
2. Aebi, M., and C. Weissmann. 1987. Precision and orderliness in splicing. *Trends Genet.* 3:102-107.
3. Beyer, A., and Y. N. Osheim. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2:754-765.
4. Bindereif, A., and M. R. Green. 1987. An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J.* 6:2415-2424.
5. Black, D. L., B. Chabot, and J. A. Steitz. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-mRNA splicing. *Cell* 42:737-750.
6. Chabot, B., D. L. Black, D. M. LeMaster, and J. A. Steitz. 1985. The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. *Science* 230:1344-1349.
7. Chabot, B., and J. A. Steitz. 1987. Multiple interactions between the splicing substrate and small ribonucleoproteins in spliceosomes. *Mol. Cell. Biol.* 7:281-293.
8. Cheng, S.-C., and J. Abelson. 1987. Spliceosome assembly in yeast. *Genes Dev.* 1:1014-1027.
9. Chu, G., and P. A. Sharp. 1981. A gene chimera of SV40 and mouse B-globin is transcribed and properly spliced. *Nature (London)* 289:378-382.
10. Dobkin, C., and A. Bank. 1985. Reversibility of IVS 2 missplicing in mutant human beta-globin gene. *J. Biol. Chem.* 260:16332-16337.
11. Eperon, I. P., J. P. Estibeiro, and I. C. Eperon. 1986. The role of nucleotide sequence in splice site selection in eukaryotic pre-messenger RNA. *Nature (London)* 324:280-282.
12. Frendewey, D., A. Kramer, and W. Keller. 1987. Different small nuclear ribonucleoprotein particles are involved in different steps of splicing complex formation. *Cold Spring Harbor Symp. Quant. Biol.* 52:287-298.
13. Frendewey, D., and W. Keller. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* 42:355-367.
14. Furdon, P. J., and R. Kole. 1986. Inhibition of splicing but not cleavage at the 5' splice site by truncating human B-globin pre-mRNA. *Proc. Natl. Acad. Sci. USA* 83:927-931.
15. Furdon, P. J., and R. Kole. 1988. The length of the downstream exon and the substitution of specific sequences affect pre-mRNA splicing in vitro. *Mol. Cell. Biol.* 8:860-866.

16. Gerke, V., and J. A. Steitz. 1986. A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of premessenger RNA. *Cell* 47:973-984.
17. Grabowski, P. J., R. A. Padgett, and P. A. Sharp. 1984. Messenger RNA splicing in vitro: an excised intervening sequence and a potential intermediate. *Cell* 37:415-427.
18. Grabowski, P. J., S. R. Seiler, and P. A. Sharp. 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* 42:345-353.
19. Grabowski, P. J., and P. A. Sharp. 1986. Affinity chromatography of splicing complexes: U2, U5 and U4/U6 small nuclear ribonucleoprotein particles in the spliceosome. *Science* 233:1294-1299.
20. Green, M. R. 1986. Pre-mRNA splicing. *Annu. Rev. Genet.* 20:671-708.
21. Hawkins, J. D. 1988. A survey on intron and exon lengths. *Nucleic Acids Res.* 16:9893-9908.
22. Hornig, H., M. Aebi, and C. Weissmann. 1986. Effect of mutations at the lariat branch acceptor site on B-globin pre-mRNA splicing in vitro. *Nature (London)* 324:589-591.
23. Konarska, M. M., R. A. Padgett, and P. A. Sharp. 1985. Trans splicing of mRNA precursors in vitro. *Cell* 42:165-171.
24. Konarska, M. M., and P. A. Sharp. 1986. Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell* 46:845-855.
25. Konarska, M. M., and P. A. Sharp. 1987. Interactions between small nuclear ribonucleoprotein particles in the formation of spliceosomes. *Cell* 49:763-774.
26. Krainer, A. R., and T. Maniatis. 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell* 42:725-736.
27. Kuhne, T., B. Wieringa, J. Reiser, and C. Weissmann. 1983. Evidence against a scanning model for RNA splicing. *EMBO J.* 2:727-733.
28. Lang, K. M., and R. A. Spritz. 1983. RNA splice site selection: evidence for a 5' to 3' scanning model. *Science* 220:1351-1355.
29. Langford, C. J., and D. Gallwitz. 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. *Cell* 33:519-527.
30. Lamond, A. I., M. M. Konarska, and P. A. Sharp. 1987. A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. *Genes Dev.* 1:532-543.
31. Lewin, B. 1980. Alternatives for splicing: recognizing the ends of introns. *Cell* 22:324-326.
32. Maniatis, T., and R. Reed. 1987. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature (London)* 325:673-678.
33. Mitchell, P. A., G. Urlaub, and L. Chasin. 1986. Spontaneous splicing mutations at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol. Cell. Biol.* 6:1926-1935.
34. Mount, S. M., I. Pettersson, M. Hinterberger, A. Karmas, and J. A. Steitz. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell* 33:509-518.
35. Munroe, S. H. 1988. Antisense RNA inhibits splicing of pre-mRNA in vitro. *EMBO J.* 7:2523-2532.
36. Naora, H., and N. J. Deacon. 1982. Relationship between the total size of exons and introns in protein-coding genes of higher eukaryotes. *Proc. Natl. Acad. Sci. USA* 79:6196-6200.
37. Nelson, K. K., and M. R. Green. 1988. Splice site selection and ribonucleoprotein complex assembly during in vitro pre-mRNA splicing. *Genes Dev.* 2:319-329.
38. Ohno, M., H. Sakamoto, and Y. Shimura. 1987. Preferential excision of the 5' proximal intron from mRNA precursors with two introns as mediated by the cap structure. *Proc. Natl. Acad. Sci. USA* 84:5187-5191.
39. Ohshima, Y., and Y. Gotoh. 1987. Signals for the selection of a splice site in pre-mRNA. *J. Mol. Biol.* 195:247-259.
40. Parent, A. S., S. Zeitlin, and A. Efstratiadis. 1987. Minimal exon sequence requirements for efficient in-vitro splicing of monointronic nuclear pre-mRNA. *J. Biol. Chem.* 262:11284-11291.
41. Parker, R., P. G. Siliciano, and C. Guthrie. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involved base-pairing to the U2-like snRNA. *Cell* 4:229-239.
42. Pikielny, C. W., and M. Rosbash. 1985. mRNA splicing efficiency in yeast and the contribution of non-conserved sequences. *Cell* 41:119-126.
43. Pikielny, C. W., B. C. Rymond, and M. Rosbash. 1986. Electrophoresis of ribonucleoproteins reveals an ordered assembly of yeast splicing complexes. *Nature (London)* 324:341-345.
44. Rautmann, G., and R. Breathnach. 1985. A role for branch points for splicing in vivo. *Nature (London)* 315:430-432.
45. Rautmann, G., H. W. D. Matthes, M. J. Gait, and R. Breathnach. 1984. Synthetic donor and acceptor sites function in an RNA polymerase B (II) transcription unit. *EMBO J.* 3:2021-2028.
46. Reed, R., and T. Maniatis. 1985. Intron sequences involved in lariat-formation during pre-mRNA splicing. *Cell* 41:95-105.
47. Reed, R., and T. Maniatis. 1986. A role for exon sequences and splice-site proximity in splice site selection. *Cell* 46:681-690.
48. Ruby, S., and J. Abelson. 1988. An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science* 242:1029-1035.
49. Ruskin, B., and M. R. Green. 1985. Specific and stable intron-factor interactions are established early during in vitro pre-mRNA splicing. *Cell* 43:131-142.
50. Ruskin, B., J. M. Greene, and M. R. Green. 1985. Cryptic branch point activation allows accurate in vitro splicing of human B-globin intron mutants. *Cell* 41:833-844.
51. Sharp, P. A. 1981. Speculations on RNA splicing. *Cell* 23:643-646.
52. Sharp, P. A. 1987. Splicing of messenger RNA precursors. *Science* 235:766-771.
53. Solnick, D. 1985. Alternative splicing caused by RNA secondary structure. *Cell* 43:667-676.
54. Somasekhar, M. B., and J. E. Mertz. 1985. Exon mutations that affect the choice of splice sites used in processing the SV40 late transcripts. *Nucleic Acids Res.* 13:5591-5609.
55. Tazi, J., C. Alibert, J. Temsamani, I. Reveillaud, G. Cathala, C. Brunel, and P. Jeanteur. 1986. A protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. *Cell* 47:755-766.
56. Treisman, R., S. Orkin, and T. Maniatis. 1983. Specific transcription and RNA splicing defects in five cloned B-thalassemia genes. *Nature (London)* 302:591-596.
57. Treisman, R., S. Orkin, and T. Maniatis. 1983. Structural and functional defects in B-thalassemia, p. 99-121. *In* G. Stamatoyannopoulos and A. Nienhuis (ed.), *Globin gene expression and hematopoietic differentiation*. Alan R. Liss, Inc., New York.
58. Treisman, R., N. J. Proudfoot, M. Shander, and T. Maniatis. 1982. A single-base change at a splice in a B<sup>0</sup>-thalassemic gene causes abnormal RNA splicing. *Cell* 29:903-911.
59. Wieringa, B., E. Hofer, and C. Weissmann. 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit B-globin intron. *Cell* 37:915-925.
60. Wieringa, B., F. Meyer, J. Reiser, and C. Weissmann. 1983. Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit B-globin gene. *Nature (London)* 301:38-43.
61. Zapp, M. L., and S. M. Berget. 1989. Evidence for nuclear factors involved in recognition of 5' splice sites. *Nucleic Acids Res.* 17:2655-2674.
62. Zhuang, Y., and A. M. Weiner. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* 46:827-835.
63. Zillmann, M., S. D. Rose, and S. M. Berget. 1987. U1 small nuclear ribonucleoproteins are required early during spliceosome assembly. *Mol. Cell. Biol.* 7:2877-2883.
64. Zillmann, M., M. L. Zapp, and S. M. Berget. 1988. Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonuclear protein particles. *Mol. Cell. Biol.* 8:814-821.